Abstract Title Page

Not included in page count.

Title: What is the Minimum Information Needed to Estimate Average Treatment Effects in Education RCTs?

Authors and Affiliations: Peter Z. Schochet, Mathematica Policy Research, Inc.

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

Randomized controlled trials (RCTs) are considered the "gold standard" for evaluating an intervention's effectiveness. While the number of RCTs conducted in the field of education has increased significantly in recent years (Puma et al. 2009), there are still many times when the lack of rigorous evidence of efficacy prevents education policymakers and administrators from making evidence-based decisions. With the help of a research team, state and district administrators can help to fill this knowledge gap by taking advantage of opportunities to introduce an experiment into the normal course of action. Not only can this approach be cost effective, these "*opportunistic experiments*" can generate strong evidence to inform educational decisions. This approach can also be used to *replicate* promising findings from previous evaluations, which is a theme of the SREE conference.

Recently, the federal government has placed increased emphasis on the use of opportunistic experiments. In July 2013, the Office of Management and Budget released guidance for 2013 agency budget submissions that encouraged agencies to propose "high-quality, low-cost evaluations" that "should help agencies improve the quality and timeliness of evaluations—for example, by building evaluation into ongoing program changes and by reducing costs by measuring key outcomes in existing administrative data sets" (Burwell et al. 2013).

A key criterion for conducting opportunistic experiments, however, is that there is relatively easy access to data about key outcomes. Administrative record data are a possible source for such outcome data. State longitudinal data systems (SLDSs) contain statewide student data that can be linked over time and, potentially, to additional data sources. Since 2006, 47 states have received at least one grant from IES' SLDS Grant Program to support the design, development, implementation, and expansion of the data systems. The key advantage of working with SLDSs is that they typically include data from all public schools in the state.

Access to these administrative records data, however, is often difficult due to state data confidentiality concerns. The federal Family Educational Rights and Privacy Act (FERPA) prohibits the disclosure of personally identifiable information from education records without written consent, except under certain exceptions. States interpret FERPA in different ways, however, to protect the confidentiality of their data. States also have their own privacy laws, which vary and in some cases are highly restrictive. Some states will release identifiable student-level data if data security and confidentiality protections are sufficient and IRB approval is granted, but others will not.

One possible approach for facilitating access to SLDS data for opportunistic experiments is to minimize the data requests for individual-level data and to instead request *aggregate* data that can be used for impact estimation. Yet there is no methods literature in the education field that addresses what aggregate statistics need to be requested to obtain rigorous estimates of average treatment effects (ATEs) for various types of designs used in education research. This paper will fill this gap.

Purpose / Objective / Research Question / Focus of Study:
Description of the focus of the research.

>This paper will address the following research question: What is the *minimum* amount of information that researchers can request from state SLDS staff to obtain unbiased estimates of ATEs and their standard errors for the full population and key population subgroups? Clearly, obtaining more disaggregated data would allow researchers to obtain a fuller range of research questions regarding intervention effects than the use of highly aggregated data (such as mediated analyses). However, our focus is on identifying aggregate statistics that can be used to address key confirmatory evaluation questions with the goal of minimizing state staff effort in processing the data and state confidentiality concerns.

The paper will consider non-clustered designs where (1) students are randomized to a treatment and control group, (2) clustered designs where units (such as schools or classrooms) are randomized, and (3) both stratified designs where random assignment is conducted within blocks (such as school districts or schools) and non-stratified designs. The paper will focus on the estimation of ATEs on a continuous student achievement test score that is available in the SLDS data. We will consider impact estimation models that adjust for baseline covariates and those that do not.

Setting:
Description of the research location.
(May not be applicable for Methods submissions)

> Not Applicable.

Population / Participants / Subjects:
Description of the participants in the study: who, how many, key features, or characteristics.
(May not be applicable for Methods submissions)

> Not applicable, except that the methods are demonstrated using data discussed below.

Intervention / Program / Practice:
Description of the intervention, program, or practice, including details of administration and duration.
(May not be applicable for Methods submissions)

> Not applicable.

Significance / Novelty of study:
Description of what is missing in previous work and the contribution the study makes.

> See above.

Statistical, Measurement, or Econometric Model:
Description of the proposed new methods or novel applications of existing methods.

> The statistical methods that we will discuss to identify minimum data requirements for impact estimation will hinge on the RCT design. We will consider standard ordinary least squares (OLS) methods for design where students are randomized to the treatment and control groups. For *clustered* RCT designs, we will not be able to use standard hierarchical linear model (HLM) methods (Bryk and Raudenbush, 2002) that are typically used by education researchers for impact estimation. This is because these methods rely on *iterative* maximum likelihood procedures to estimate the variance components that do not accommodate aggregate data.

Instead, for clustered designs, we will use *design-based methods* where the data are averaged to the cluster level (for example, school level). These include the estimators by Schochet (2013) and Baltagi and Chang (1994). These types of design-based approaches are not often used by education researchers, but are often used in other disciplines, and have the advantage over HLM methods that the weighting scheme used to weight individual clusters to obtain ATEs is more transparent than for HLM methods. Similar design-based methods will be used for stratified designs where random assignment is conducted within blocks and block effects are treated as random. These designs will include RCTs with pairwise matching of schools.

A contribution of this article is that it will use a unified framework for developing estimators, variance formulas, and approaches for significance testing that education researchers can apply using the aggregate statistics from the administrative records. Model assumptions will be carefully spelled out, such as whether the estimators are assumed to generalize to a broader population or to the fixed study sample only.

Usefulness / Applicability of Method:
Demonstration of the usefulness of the proposed methods using hypothetical or real data.

> The article will provide examples of the new methods using several published RCTs that were funded by the Institute of Education Sciences (IES) at the U.S. Department of Education (ED), the Department of Labor, and several foundations. These RCTs tested the effects of a wide range of education interventions, including early elementary school math curricula, , selected reading comprehension interventions, Teach for America, and Job Corps. Across the RCTs, random assignment was conducted at either the student level or the school or teacher (classroom) level in low-performing school districts, and for most studies, the key outcome measures were math or reading test scores of elementary school students.

Research Design:
Description of the research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).
(May not be applicable for Methods submissions)

>Not applicable.

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.

(May not be applicable for Methods submissions)

> Not applicable.

Findings / Results:
Description of the main findings with specific details.
(May not be applicable for Methods submissions)
> Not applicable.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

> There is increasing interest in policy circles for educators to conduct low-cost opportunistic experiments to build the evidence base for identifying promising interventions, and to replicate findings from previous evaluations in alternative contexts. Such RCTs, however, will be feasible only if administrative records data are readily available to conduct such analyses. Yet, obtaining these data can be difficult and time-consuming due to data access issues and staff availability.

The purpose of this paper will be to examine the minimum amount of SLDS information that researchers can request to rigorously address confirmatory analysis questions for a full range of RCT designs used in education research. The paper will demonstrate the use of these methods using several real-world examples of education RCTs.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Baltagi, B. and Y. Chang (1994). A Comparative Study of Alternative Estimators for the Unbalanced One-Way Error Component Regression Model. Journal of Econometrics 62, 67-89.

Bryk, A. and S. Raudenbush (1992). Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park, CA: Sage.

Burwell, S., Muñoz, C., Holdren, J., and Krueger, A. "Next Steps in the Evidence and Innovation Agenda." Memorandum to the Heads of Departments and Agencies, Office of Management and Budget, Executive Office of the President, July 26, 2013. Retrieved from http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-17.pdf

Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price (2009). *What to Do When Data Are Missing in Group Randomized Controlled Trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Schochet, Peter Z. "Estimators for Clustered Education RCTs Using the Neyman Model for Causal Inference." *Journal of Educational and Behavioral Statistics*, vol. 38, no. 3, June 2013, pp 219-238.